图书情报工作
第62卷 第16期 2018年8月

基于"公平性测试"的Altmetrics学术质量评价方法研究*

■ 刘红煦[1] 王铮[2,3]
[1] 同济大学图书馆 上海 200092 [2] 西北大学公共管理学院 西安 710127
[3] 南京大学信息管理学院 南京 210023

摘要:［目的/意义］为了探讨在学科论文评价上,Altmetrics指标间的量度结果是否具有一致性,试图通过已有测量数据构建基于Altmetrics的文献综合评价模型,为评价数据的使用和重用创建新的度量标准。［**方法/过程**］获取PLoS ALMs上土木工程学科2009年至2016年发表的全部文章的Altmetrics指标,采用相关分析、主成分分析方法将15个原始指标划分到4个主成分,得到适用于具体学科的Altmetrics评价指标体系,并对Altmetrics的单一指标与综合评价指标进行相关度分析,采用"公平性测试"方法探索时间对论文评价相关度的差异。［**结果/结论**］研究发现采用"公平性测试"方法消除时间影响后,引用指标相关性上升,而与社交媒体传播相关的指标有所下降,但Twitter指标不降反增,且始终与F值具有较高的相关性。

关键词: 替代计量学 主成分分析 公平性测试 学科评价
分类号: G250
DOI: 10.13266/j.issn.0252-3116.2018.16.012

1 引言

2010年10月,J. Priem联同D. Taraborelli等发表了"Altmetrics:A manifesto"宣言,正式提出了"Altmetrics"(Alternative Metrics)术语[1]。2012年至今,国内图情学科和期刊出版领域对Altmetrics做了理论和实证探索。国内学者将其译为"替代计量学""补充计量学""社媒影响计量学"等,但该翻译在国内语境中有歧义,如J. Liu认为Altmetrics旨在取缔单一使用引文计量方法进行研究影响力评价,进入研究影响力评价的多元化、多样化阶段,绝不是取代引文计量方法[2],因此本文仍然采用可靠而简单的原词。2013年,美国国家信息标准协会宣布了一项新的两阶段项目,研究并开发基于社区的补充计量标准或是推荐做法[3]。2016年,ProQuest子公司Ex Libris宣布已将Altmetrics集成到Summon发现服务之中[4],极大程度上改善了用户体验,促进内容的发现与集成。Altmetrics有助于填补引用指标无法表达的知识缺口,Altmetrics不仅在数据的种类上比传统文献计量学更多,而且在数据的量级上更庞大。尽管对Altmetrics的理论研究与实践仍处于起步探索阶段,存在一些前瞻性问题值得探索[5-6],但Altmetrics通过其"大众评审"[7]的优势,在一定程度上弥补了同行评议的不足。在评价单篇论文方法优越性上,Altmetrics指标在设立之初与传统计量学中引用指标并不存在可比性,一是两者体现出文献的不同方面影响力,二是Altmetrics指标与传统计量学中引用指标更多地体现出一种互补关系,Altmetrics评价工具ImpactStory与Plum Analytics都将传统计量学中的被引量作为影响力评估来源。2017年3月,Altmetric.com将Web of Science中的被引次数也作为指标之一,通过与传统计量学中引用指标结合,Altmetrics指标已经更加精准地评估文献影响力。

目前的研究大多聚焦在验证Altmetrics指标与其他文献计量指标量度结果相关性,通过检验单一Altmetrics指标优于综合Altmetrics指标评价学术论文影响力,如通过对所选样本进行同行评议或者其他第三方指标进行评价验证,目前相关的研究已经验证了

* 本文系同济大学图书馆2017年馆内项目"替代计量学与学术评价体系变革"研究成果之一。
作者简介:刘红煦(ORCID:0000-0002-4809-1437),助理馆员,硕士,E-mail:hxliu@lib.tongji.edu.cn;王铮(ORCID:0000-0001-5727-5935),讲师,博士后。
收稿日期:2018-02-01 修回日期:2018-05-24 本文起止页码:102-110 本文责任编辑:徐健

Altmetrics 不是孤立的,在未来具有一定的应用潜力。本研究则定位在 Altmetrics 指标间的量度结果的一致性研究,采用主成分分析、降维等操作,通过已有测量数据构建基于 Altmetrics 的文献综合评价模型,为评价数据的使用和重用创建新的度量标准。

在确定较高的测量值是否与较高的文章某一具体指标相关联的过程中,需要设计一个公正的测试。究其原因是因为相关性测试对于 Altmetrics 来说并不理想,许多测试都基于快速增长的数据,平均来说,较新的文章可以比较旧的文章获得更高的比较分数。指标数据值也需要时间累积,如果不调整这些差异,则相关性检验总是偏向于负相关。通过调整引用或使用数据以消除这些偏见(如下载统计数据[6－7])是很困难的,因为可靠的"使用数据"仅适用于发表时间最近的引用窗口所使用的文献。为了避免这些问题,可以采用"公平性测试"方法消除时间差。

2 国内外相关研究进展

2.1 Altmetrics 与传统指标之间的相关性探索

为了验证 Altmetrics 与传统指标之间的相关性,已经有研究人员进行了多角度的尝试,研究结果不一。针对传统引用指标与 Altmetrics 指标的相关强度,K. Yan 等[8]发现 HTML 浏览量和 PDF 下载量遵循长尾分布规律,两者的相关程度较强;X. Wang 等[9]发现被引量与 PDF 浏览量之间的相关程度显著高于其他指标。而针对 Altmetrics 指标内部的相关性,H. Alhoori 等[10]发现 Altmetrics 之间的相关性不强,且单篇论文在不同平台上的被关注度也不同。此外,Altmetrics 与引用指标的关系收到了广泛关注,L. Bornmann 等[11]发现博客、微博与引用指标的相关程度较低,而书签的表现稍好一些;R. Costas 等[12]发现不同学科领域存在差异性的分布规律,并且只适用于最新出版的论文;A. Zuccala 等[13]发现在论著中被引量与读者评论数之间也不具有显著相关性;K. Chen 等[14]验证了不同学科、不同地域学者的引用与 Altmetrics 相关度不强。M. Thelwall 等[15]通过具体模型进行检验,也表明 Altmetrics 与引用指标存在不同强度的正相关关系。J. Priem[16]认为单一指标用于学术评价仍存在缺陷,必须将各个指标结合起来,但并不是所有的替代计量指标都能达到良好的效果;J. Lin[17]提出将替代计量和同行评议相结合的评价体系,但后续使用收效甚微;我国学者余厚强也将网络环境下学术评价的结果外化为"投入程度"与"利用深度"[18],但是也没有给出具体的衡量标准。

2.2 基于 PLOS 的论文级别计量研究现状

论文层面计量学[19－20]与 Altmetrics[21－22]不是同义词,二者的内涵与研究范围不同,前者是试图在论文层面测量个体论文的影响力。在测量过程中,论文层面计量学搜集不同的数据源,而后者则是利用新型数据源测量多种对象(如一篇论文、一本期刊或者一个独立的学者)的影响力。论文层面计量学是利用 Altmetrics 数据源与传统数据源,在论文层面重新定义论文的影响力。Altmetrics 则是深入研究各种新型数据源的特点、规律与应用,不仅仅是在论文层面对各种数据源进行聚合[23]。

早在 2009 年,C. Neylon 和 S. Wu[24]就曾以 PLoS 和 F1000 为例,对论文层级的学术影响力计量(ALMs)方案的可行性进行探讨,并分别从计量数据的来源和专家评论的激励机制两个角度予以论述。PLOS 的 Altmetrics 计量指标主要针对 PLOS 出版的文献,其 ALMs 指标主要分为 5 大类[25]:①浏览(VIEWED),含 PLOS Journals(HTML、PDF、XML)、PubMed Central(HTML、PDF)指标;②被引用(CITED),含 CrossRef、Datacite、Europe PMC、PubMed Central、Scopus、Web of Science 指标;③被保存(SAVED),含 CiteULike、Mendeley 指标;④被讨论(DISCUSSED),含 PLOS Comments、Facebook、Reddit、Twitter、Wikipedia 指标;⑤被推荐(RECOMMENDED),含 F1000 Prime 指标。虽然 PLOS 的计量指标相对较少,但是每个计量指标的数据来源渠道比较权威且被广泛使用[26]。王睿发现[27] PLoS ALMs 是最早提出利用 Altmetrics 指标进行单篇论文影响力评价,并成功将这一理论付诸实施的工具。刘晓娟等[28]基于 PLOS ALM 数据,发现不同维度的文献价值不同,评价工作中需要根据指标自身的优势、劣势和适用范围,合理构建指标体系。

2.3 Altmetrics 在学科评价及科研服务中的应用

Altmetrics 指标与传统的引文分析并不是简单的相关与否,不同学科之间的差异直接影响了其相关性[29－30],此外,利用 Altmetrics 工具可以为学者提供学术履历服务,通过提供相应的评估服务,帮助学者塑造更加多彩的职业形象[31]。部分高校已经尝试结合 Altmetrics 提供多样化的学科服务。匹兹堡大学图书馆系统利用 Altmetrics 工具开展大学研究成果评估试点工作,与研究人员建立更为直接的沟通机制,形成多方参与的影响力评价方案[32]。此外,麻省理工学院医学院图书馆通过练习并使用 Altmetrics 产品促进 Altmetric.com 在馆内机构库中的具体化应用;伦敦政治经济学

院研究在线机构库与图书馆研究支持服务部门协同合作,利用 Altmetrics 为研究人员展现社会影响图景,帮助研究人员追踪他们工作的在线关注动态[33]。

2.4 “公平性测试”方法在论文影响力评价中的应用

M. Thelwall 认为[15]进行 Altmetrics 指标和引文指标相关分析的前提是消除时间窗口对数据采样的影响,提出将待测数据置于一个相对公平的测试环境中,从而提出“公平性测试”(fairness test)方法。指标分析的根本问题是各个指标数据存在偏差,以引文为例,不同学科的论文在同一时间获得的引文量有时在数量级上存在差异,通过改用适当修正的指标可以消除学科间的这种偏见(如 CNCI),而若不进行适当的处理,这种偏见可能会使比较变得毫无意义。针对学科间的差异的这种操作属于“公平性测试”的处理范畴,但针对同一学科在不同时间发表的论文的影响力测度时,也需要消除时间的影响,使得在历时性的指标考察时可靠性更高。J. Priem、H. Piwowar[16]检查了 PLoS 文章随时间推移的社交媒体事件的分布情况,注意到其中的行为差异。例如,引文、网页浏览量和维基百科引用量随着时间推移而增加,而 CiteULike、Mendeley、Delicious 书签和 F1000 等级则相对不受文章发表时期的影响。其他指标则差异巨大,是由于数据服务的变化及局限性阻碍了分析,这突出表明需要考虑到某些指标的稳定性以及在纵向研究中使用这些指标的问题。研究发现多数学者只对其所在领域最新的文章感兴趣。因此,笔者推测一篇文章在其发布日受到关注最多,随着时间关注度逐渐下降。

具体来说,在“公平性测试”中,每篇文章只与紧接在它之前和之后发布的两篇文章(同一期刊)进行比较。因此,只有发布时间大致相同的文献,最大限度地降低了引用延迟和使用偏差,彼此才可进行比较。同时,在被测试文章之后发布的文章的任何微弱的优势或劣势都应该通过其平均值与之前发布的文章的等价优势或劣势进行抵消。该测试给出了 3 种可能的结果:①成功:分数评分高于相邻两篇文章的平均分,且其引文评分高于两篇相邻文章的平均值,或者评分低于相邻两篇文章的平均分,且分数低于两篇相邻文章的平均值;②失败:与“成功”条件相反,即分数评分高于相邻两篇文章的平均分,且其引文评分低于两篇相邻文章的平均值,或是分数评分低于相邻两篇文章的平均分,且其引文评分高于两篇相邻文章的平均值;③无效:所有其他情况。对于每个比较对,都会创建一个列表。因此,测量得分为正的文章通过该测量有效收集了数据。由于数据收集过程因信息量的不同而不同,因此不可能假设一个比对的文章的正分数意味着它也将被有效地监测其他计量学指标。以引用指标为例,尽管由于引文时间窗口较短,文章的引用得分不可靠,但这不应该导致结果出现系统性偏差,因为在用某一指标与对照得分进行比较的方法中考虑了发表时间。

F. Radicchi 等[34-35]采用“公平性测试”方法,从概率的角度为学术期刊评价指标的学科间比较的公平性提供可计量的评价依据。两人利用“公平性测试”方法,实证分析在学科的 top 10% 论文中,归一化指标 fractional citation 的公平性要远远高于期刊影响因子。陈福佑等[36]以学科归一化的期刊评价指标相对期刊影响因子和平均百分比等级指标为例,引入“公平性测试”方法检验了两个学科归一化评价指标在不同学科比较中的适用性。王睿等[37]以 273 篇高替代计量指标论文为样本,利用“公平性测试”方法消除时间对被引次数的影响,发现具有高替代计量指标的学术论文同时具有较高的被引次数。

2.5 小结

通过对 Altmetrics 的相关研究进行分析,发现学者们不断探索 Altmetrics 的内涵与应用,但在 Altmetrics 效度实证方面,仍有待进一步深入研究。例如:由于许多因素,如论文的发表时间、所属学科、期刊、引用动机、地域及数据的丰富性都可能造成实验结果的偏差,这方面的分组研究却很少,因此笔者选取一个具体的学科,选择某一期刊,探讨学科论文的评价模式。为了探讨在学科论文评价上,结合论文引用指标、论文使用指标和论文 Altmetrics 指标是否优于单一使用一种类型指标,且评价是否具有一定可信度问题,笔者提出了 Altmetrics 在具体学科领域论文评价中的应用探析这一研究主题。事实上,不同学科研究人员工具使用偏好存在较大差异,当前的 Altmetrics 测量方法更多的是适用于 STEM(科学、技术、工程与数学)领域,而其他学科(人文、艺术与社会科学)没有多少可用的工具和指标,因此笔者选择工程学中的一个具体学科——土木工程作为分析的案例,探讨 Altmetrics 指标在该学科论文评价中的适用性。此外,采用“公平性测试”方法以消除发表时间对期刊论文影响力评价的影响,并对这一方法进行简化,最大程度地提高分析的可靠性。

3 研究设计

为了探讨 Altmetrics 指标间量度结果的一致性,笔

者将采用相关分析、主成分分析降维等操作,通过已有测量数据构建单一学科基于 Altmetrics 的文献综合评价模型,为评价数据的使用和重用创建新的度量标准。本研究首先获取土木学科一段时期内文献的 Altmetrics 指标,根据数据库所收录的指标类型,获取浏览(viewed)、保存(saved)、讨论(discussed)、引用(cited)4 大类 15 个细分指标,通过主成分分析构建基于 Altmetrics 的文献综合评价模型,得到综合 F 计算公式,比较各 Altmetrics 指标与 F 的相关性,采用"公平性测试"方法消除时间影响后,重新比较相关性并探讨哪些因素受到时间的影响较大,且在未来基于 Altmetrics 的论文社交影响力指标评价体系中,需重新划分相关因素的权重。

3.1 土木学科论文样本选取

考虑到论文从发表到浏览、保存、讨论乃至引用需要一定的时间,因此在 PLoS ALMs(https://www.plos.org/article-level-metrics)上检索 2009 年至 2016 年出版的主题为 Civil Engineering 的全部文章,检索时间为 2017 年 11 月 3 日,类型选择 Research Article,共检索到 410 篇文献,下载这些文献的 Metrics Data 数据。通过初步查看数据发现文献的 Web of Science 引用次数均为 0,作者推测这与 PLoS 数据库收录范围有关[38],故将全部文献以 DOI 标识符在 Web of Science 核心合集中进行匹配,获得每篇论文的 Web of Science 最近 180 天使用次数、Web of Science 从 2013 年至今使用次数、Web of Science 至今引用次数等指标。此外,Metrics Data 表中数据 90% 以上为 0,不具有分析价值,故删除这部分字段。最终纳入分析的数据示意及基本分布情况(见表 1)。其中代码列用于指标体系构建,WOS Short Use 指文章在 Web of Science 上最近 180 天使用次数,WOS long Use 指文章在 Web of Science 上 2013 年至今使用次数。统计分析软件采用 StataSE12。

表 1 基于 PLOS 的土木工程学科 ALMS 计量指标

类别	指标	指标缩写	代码	均值	标准差	最小值	最大值
viewed	PLOS views	PLOSvi	V1	3 160.932	3 471.857	376	25 421
	PLOS PDF downloads	PLOSPDFd	V2	485.422	498.855 2	96	4 855
	PLOS XML downloads	PLOSXMLd	V3	40.926 83	67.502 78	2	504
	PMC views	PMCviews	V4	441.112 2	482.194 9	26	4 400
	PMC PDF Downloads	PMCPDFD	V5	106.478	95.492 27	7	615
saved	Mendeley	Mendeley	S1	39.2	48.214 47	0	393
	Figshare	Figshare	S2	74.441 46	112.177 3	0	1 508
	WOS Short Use	WOSSU	S3	2.619 512	2.876 702	0	19
	WOS long Use	WOSlU	S4	28.204 88	29.489 08	0	253
discussed	Facebook	Facebook	D1	18.668 29	218.760 6	0	4 392
	Twitter	Twitter	D2	7.092 683	15.367 58	0	178
	Wikipedia	Wikipe	D3	0.095 122	0.440 3	0	4
cited	Scopus	Scopus	C1	10.370 73	22.997 48	0	347
	PMC Europe Citations	PMCEuCt	C2	1.680 488	4.634 606	0	45
	Web of Science	WOSCit	C3	8.758 537	18.794 29	0	273

3.2 综合得分计算方式

关于 Altmetrics 的研究越来越受到关注,国内外学者也提出相关评价指标,并通过大量的实证研究验证这些指标的可行性。Altmetrics 在计算综合分数时,通常需要根据不同社会化媒体影响力的不同为其赋予适当的权重。2017 年 9 月,www.altmetric.com 给出的一般计算方式见表 2[39]。

根据 Altmetric 评价系统的评分标准来看,Altmetric 评价系统更侧重于大众和网络影响力,但是并未把下载和阅读次数考虑在内。结合 www.altmetric.com 对 Altmetric 的计算方法及 Plos 本身的计算方法,通过

表 2 www.altmetric.com 不同数据源赋分情况

数据来源	分值	数据来源	分值
News	8	Wikipedia	3
Blogs	5	Policy Documents(per source)	
Open Syllabus	1	LinkedIn	0.5
Google +		YouTube	0.25
Twitter		Reddit/Pinterest	
Sina Weibo		Q&A	
F1000/Publons/Pubpeer		Facebook	

对指标数据源进行统计和计算,设计出适用于具体学科学术环境的 Altmetric 公式,即土木工程学科结合 Altmetrics 的文献计量算法。

3.2.1 主成分分析 各指标主成分分析的结果是否可行通过对 KMO(Kaiser-Meyer-Olkin)和 SMC(squared multiple correlations)进行进一步检验(见表 3)。KMO 抽样充分性测度也是用于测量变量之间相关关系的强弱的重要指标,是通过比较两个变量的相关系数与偏相关系数得到的。可以看出 KMC 检验结果为 0.8025,主成分分析能够起到比较好的数据约化效果。SMC 即一个变量与其他所有变量的复相关系数的平方,也就是复回归方程的可决系数,样本 SMC 较高,可以看出变量的线性关系较强,可以做主成分分析。

表 3 KMO 和 SMC 检验可行性结果

变量	KMO 检验	SMC 检验	变量	KMO 检验	SMC 检验
PLOSvi	0.8218	0.7805	WOSIU	0.7606	0.545
PLOSPDFd	0.8493	0.8748	Facebook	0.5747	0.3886
PLOSXMLd	0.6584	0.1895	Twitter	0.6819	0.5017
PMCviews	0.7463	0.5088	Wikipe	0.7541	0.1422
PMCPDFD	0.7743	0.6148	Scopus	0.7782	0.9662
Mendeley	0.8861	0.8306	PMCEuCt	0.9212	0.7068
Figshare	0.6466	0.0735	WOSCit	0.7876	0.9671
WOSSU	0.5944	0.264	Overall	0.8025	

前四个成分特征值累积占了总方差的 67.743%。后面的特征值的贡献越来越少。根据主成分个数确定准则,取特征值大于 1 的前四个主成分,其因子累计方差贡献率接近 80%。因此,将 15 个原始指标划分到前四个主成分中,见表 4。通过对载荷矩阵进行旋转,可得到相应的特征向量,见表 5。

表 4 指标方差计算结果

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.60161	3.78558	0.3734	0.3734
Comp2	1.81603	0.325359	0.1211	0.4945
Comp3	1.49067	0.237554	0.0994	0.5939
Comp4	1.25312	0.264345	0.0835	0.6774
Comp5	0.988773	0.075761	0.0659	0.7433
Comp6	0.913012	0.02113	0.0609	0.8042
Comp7	0.891883	0.361494	0.0595	0.8637
Comp8	0.530389	0.113876	0.0354	0.899
Comp9	0.416513	0.061145	0.0278	0.9268
Comp10	0.355368	0.055323	0.0237	0.9505
Comp11	0.300046	0.099031	0.02	0.9705
Comp12	0.201014	0.068984	0.0134	0.9839
Comp13	0.132031	0.039773	0.0088	0.9927
Comp14	0.092258	0.07498	0.0062	0.9988
Comp15	0.017278		0.0012	1

3.2.2 确定指标体系 在主成分分析中,四个主成分下的数值代表相互关系的紧密程度。从表 5 主成分与各变量载荷矩阵可知,PLOSvi、PLOSPDFd、PLOSXMLd、PMCPDFD、Mendeley、Scopus、PMCEuCt、WOSCit 在第一个主成分上有较高载荷,说明第一个主成分基本能够反映这些计量指标的数据信息;Facebook、Twitter 在第二个主成分上有较高载荷;WOSSU、WOSIU 在第三个主成分上有较高载荷,说明第三个主成分基本反映了 Web of Science 上指标的信息;MCviews、Figshare、Wikipe 在第四个主成分上有较高载荷。总之,从中提取四个主成分可以基本反映全部指标的信息,因此,笔者决定用四个主成分来代替原来的 15 个指标。

表 5 主成分分析后成分结构

变量	Comp1	Comp2	Comp3	Comp4
PLOSvi	0.779	0.427	-0.055	-0.01
PLOSPDFd	0.925	0.049	0.043	-0.121
PLOSXMLd	0.296	0.035	-0.293	-0.176
PMCviews	0.522	0.009	-0.379	0.559
PMCPDFD	0.629	-0.128	-0.424	0.358
Mendeley	0.892	-0.056	0.242	0.024
Figshare	0.167	0.059	-0.12	0.421
WOSSU	0.165	0.163	0.769	0.175
WOSIU	0.528	0.093	0.588	0.309
Facebook	0.083	0.788	-0.131	-0.216
Twitter	0.249	0.841	-0.109	-0.164
Wikipe	0.272	0.167	-0.028	0.41
Scopus	0.861	-0.265	0.052	-0.296
PMCEuCt	0.777	-0.288	-0.175	-0.246
WOSCit	0.875	-0.248	0.065	-0.27

将 4 个主成分因子分别记为 F1、F2、F3、F4,则前四个主成分的表达式是:

$$F1 = 0.779 * V1 + 0.925 * V2 + 0.296 * V3 + 0.522 * V4 + 0.629 * V5 + 0.892 * S1 + 0.167 * S2 + 0.165 * S3 + 0.528 * S4 + 0.083 * D1 + 0.249 * D2 + 0.272 * D3 + 0.861 * C1 + 0.777 * C2 + 0.875 * C3$$

$$F2 = 0.427 * V1 + 0.049 * V2 + 0.035 * V3 + 0.009 * V4 - 0.128 * V5 - 0.056 * S1 + 0.059 * S2 + 0.163 * S3 + 0.093 * S4 + 0.788 * D1 + 0.841 * D2 + 0.167 * D3 - 0.265 * C1 - 0.288 * C2 - 0.248 * C3$$

$$F3 = -0.055 * V1 + 0.043 * V2 - 0.293 * V3 - 0.379 * V4 - 0.424 * V5 + 0.242 * S1 - 0.12 * S2 + 0.769 * S3 + 0.588 * S4 - 0.131 * D1 - 0.109 * D2 - 0.028 * D3 + 0.052 * C1 - 0.175 * C2 + 0.065 * C3$$

刘红煦，王铮. 基于"公平性测试"的 Altmetrics 学术质量评价方法研究[J]. 图书情报工作,2018,62(16):102－110.

F4 = －0.01 * V1 －0.121 * V2 －0.176 * V3 + 0.559 * V4 +0.358 * V5 +0.024 * S1 +0.421 * S2 + 0.175 * S3 +0.309 * S4 －0.216 * D1 －0.164 * D2 + 0.41 * D3 －0.296 * C1 －0.246 * C2 －0.27 * C3

根据以上四个主成分的表达式和四个主成分的特征向量值,可以进一步推算出评分模型。在主成分分析中,每个主成分的方差贡献率代表了该主成分所包含的信息多少,可以将每个主成分的方差贡献率占累加贡献率的比重作为该主成分的权重。由表 4 可以计算出四个主成分的权重分别为 0.551 2、0.178 8、0.146 7、0.123 3。对四个主成分求加权平均数即为学术论文综合影响力:

F =0.5512F1 +0.1788F2 +0.1467 * F3 +0.1233 * F4。

由此,可计算出各个期刊的综合得分。

4 基于"公平性测试"方法的论文学术影响力分析

4.1 各指标与 Altmetrics 综合得分相关性分析

通过学术论文综合影响力计算公式 F 计算得到全部文献综合影响力。计算各指标与 Altmetrics 综合得分相关性及在 1% 的显著性水平上进行检验的结果,得到如表 6 所示:

表 6 各指标与综合得分直接相关性及显著性水平

指标	相关性	检验结果	指标	相关性	检验结果
PLOSvi	0.992 6 *	0.000 0	WOSlU	0.416 3 *	0.000 0
PLOSPDFd	0.846 7 *	0.000 0	Facebook	0.291 9 *	0.000 0
PLOSXMLd	0.237 0 *	0.000 1	Twitter	0.482 6 *	0.000 0
PMCviews	0.467 3 *	0.000 0	Wikipe	0.282 0 *	0.000 0
PMCPDFD	0.489 7 *	0.000 0	Scopus	0.570 5 *	0.000 0
Mendeley	0.712 2 *	0.000 0	PMCEuCt	0.506 4 *	0.000 0
Figshare	0.135 8	0.507 4	WOSCit	0.592 6 *	0.000 0
WOSSU	0.110 7	0.951 9			

经过计算,Altmetrics 指标与 Scopus 引用指标、PMC Europe Citations 指标、WOS 引用指标的相关性均在 0.5 －0.6 之间,可见单一的 Altmetrics 指标与传统计量学以引用为主的各指标之间仅具有弱相关关系,构建基于多源数据指标的 Altmetrics 评价体系是解决这一问题的思路之一,如开放评审的数据、文献管理工具中的关注量与阅读量数据、社交网络中的推荐量与点击量数据等。

尽管目前文章被引用次数仍然是公认的判断文章学术影响力的重要指标,但由相关性分析结果可以看出,Views 指标中的 PLOS views 指标与综合得分相关性达到了 99.26% 。因此,如果其他计量指标和 PLOS views 指标存在相关性,说明该指标可能与学术影响力相关联。然而,将一篇论文的 Altmetrics 指标和 PLOS views 指标简单进行相关分析并不科学,由于社交网络信息传播具有短时期爆发式传播等特性,Twitter 上的信息传播往往会随时间增长急速消退,这就导致了以社交网络信息为主要信息源的 Altmetrics 指标增量随时间增长而呈现下滑的态势。此外,同时期发表的论文,往往介绍新兴前沿的文章会获得更高的 Altmetrics 指标。引文指标却与此相反,一篇论文被引用次数随时间的增长而不断增加,引用数量的变化则无显著发展态势。因此,需要对这些差异进行调整,为了结合多样化数据给予文献更精准的评价,笔者对计算方式进行改进。

4.2 采用"公平性测试"方法消除时间影响

针对 PLOS 土木学科的文献,按照综合得分计算方式,将计算出的 Altmetrics 分数和各原始指标作为变量,进行相关分析,结果相关性并不高。由此可见,需要设计实验,以验证 Altmetrics 指标和浏览数之间是否具有相关性。由于本研究论文均来自 PLOS,且属于同一学科,在每期上的分布较密集,因此选取 3 篇论文作为一组比较对象,其中 3 篇论文的前后 2 篇作为参照,选取规则是全部研究对象中发表时间距离最近的前后两篇文章,大大减少了工作量。采用新方法既能够消除时间窗口的影响,确保同一组对象的数据采样区间一致,又能保证每一组数据计算的独立性,同时减少计算成本。

采用"公平性测试"方法,以 PLOS 发表的 410 篇土木工程学科文献作为样本,查找和这 410 篇论文在样本中发表前后两篇论文作为参照。由于发表时间最早和最晚的两篇文章不具有完整的前后对比项,故最终得到 408 组比较对象,对这 408 组结果分组进行计算、比较。

4.3 结果分析

设单篇论文 Altmetrics 综合得分为 A,单篇论文某一指标为 X,研究样本中每篇文章用 p 表示,参照论文为 m、n,且 m 优先于 n 发表。将比较结果分为三组,如表 7 所示:

表 7 "公平性测试"方法规则

组别	结果一	结果二
成功组	Ap > 1/2(Am + An)且 Xp > 1/2(Xm + Xn)	Ap < 1/2(Am + An)且 Xp < 1/2(Xm + Xn)
失败组	Ap > 1/2(Am + An)且 Xp < 1/2(Xm + Xn)	Ap < 1/2(Am + An)且 Xp > 1/2(Xm + Xn)
无效组	除前两组以外情况	

表 8 "公平性测试"分析结果及排序变化

类别	比较项目	成功次数	失败次数	成功比例	失败比例	测试前排序	测试后排序	排序变化
Viewed	PLOS views	395	13	96.81%	3.19%	1	1	—
Viewed	PLOS PDF downloads	338	70	82.84%	17.16%	2	2	—
Viewed	PMC views	294	114	72.06%	27.94%	8	3	↑5
Discussed	Twitter	287	121	70.34%	29.66%	7	4	↑3
Saved	Mendeley	283	125	69.36%	30.64%	3	5	↓2
Cited	Scopus	282	126	69.12%	30.88%	4	6	↓2
Viewed	PMC PDF Downloads	270	138	66.18%	33.82%	6	7	↓1
Cited	Web of Science	262	146	64.22%	35.78%	13	8	↑5
Saved	WOS long Use	261	147	63.97%	36.03%	9	9	—
Viewed	PLOS XML downloads	258	150	63.24%	36.76%	12	10	↑2
Saved	WOS Short Use	230	178	56.37%	43.63%	15	11	↑4
Saved	Figshare	177	231	43.38%	56.62%	14	12	↑2
Discussed	Facebook	167	241	40.93%	59.07%	10	13	↓3
Cited	PMC Europe Citations	154	254	37.75%	62.25%	5	14	↓9
Discussed	Wikipedia	41	367	10.05%	89.95%	11	15	↓4

由于 Altmetrics 综合得分保留了后续三位小数，降低了比较无效的结果，通过最终比较，发现各组均不存在无效的比较结果。通过表 8 中的 Altmetric 分数以及各底层数据比较结果可以看出，成功次数远远高于失败次数。其中 PLOS views 成功率最高，达到 96.81%，这一结果验证了前文的相关性分析结果。其余测试项目的计算结果较为稳定，成功率大多在 56% –82% 这一区间。

通过对比表 6 和表 8，笔者发现"公平性测试"方法修正后，各指标与综合得分直接相关性存在差异，修正前相关性高于 0.5 的指标由高至底依次为 PLOSsvi – PLOSPDFd – Mendeley – WOSCit – Scopus – PMCEuct，修正后相关性由高至低依次为 PLOS views – PLOS PDFdownloads – PMC views – Twitter – Mendeley – Scopus – PMC PDF Downloads，消除时间窗口的干扰后，与综合指标相关度较低的指标（如 Facebook、PMC Europe Citations、Wikipedia）下降最多，而与综合指标相关度较高的指标（如 PMC views、Twitter）则上升较多，可见消除时间影响后在一定程度上拉大了指标与 F 值的差距。在测试后相关性上升的指标中，浏览指标中的 PMC views 上升最多，这与论文只有通过被浏览才能获得高的关注度有关，在 Web of Science 中的引用指标相关度上升，可以推测传统以引用为主要影响力评价指标存在一定意义，论文能够被引用，足以说明其关注度和重视程度，保存指标中的 WOS Short Use 有所上升，可见论文短期使用量也是评价论文影响力的重要指标，消除时间因素后，这一指标的重要性凸显出来。

而与社交媒体传播相关的指标在测试后相关度有所下降（如 Facebook、Wikipedia），这与其爆发式的传播特点密不可分，而这些指标普遍与 F 值相关度较低，但是值得注意的是讨论指标中的 Twitter 指标不降反增，且始终与 F 值具有较高的相关性，推测用户对 Twitter 的粘性要低于 Facebook，Twitter 的传播度更广且更自由，在时间上也呈现短时间爆发态势，Facebook 由于是好友之间的传播，其传播的稳定性更高，时间的突变性更低。

统计每一类结果的累计值，能够做如下推论：如果 Altmetrics 分数与某一指标的关系越强，则结果中成功次数与失败次数之间的差异越明显；反之，如果 Altmetrics 分数与某一指标没有关系，那么成功次数和失败次数之间没有明显差异。

5 结论与讨论

本文的创新性主要体现在以下两点：①基于 PLOS 收录的土木工程学科论文进行 ALMs 指标的主成分分析构建综合评价指标模型；②尝试应用"公平性测试"方法检测单一 Altmetrics 指标与综合 Altmetrics 得分之间相关性，并发现引入"公平性测试"方法，消除时间窗口的干扰之后，测试后相关性上升的指标中，浏览指标中的 PMC views 上升最多，这与论文只有通过被浏

览才能获得高的关注度有关。在 Web of Science 中的引用指标相关度上升,可以推测传统以引用为主要影响力评价指标存在一定意义,保存指标中的 WOS Short Use 有所上升,可见论文短期使用量也是评价论文影响力的重要指标,消除时间因素后,这一指标的重要性凸显出来。而与社交媒体传播相关的指标在测试后相关度有所下降(如 Facebook、Wikipedia),这与其爆发式的传播特点密不可分,而这些指标普遍与 F 值相关度较低,但是值得注意的是讨论指标中的 Twitter 指标不降反增,且始终与 F 值具有较高的相关性。

本研究目前仍存在一定的局限性:①PLoS ALMs 只能处理 PLoS 出版以及存储在 PubMed Central 上的论文,制约了它的应用范围;②现有自然科学领域更侧重期刊论文,而人文社会科学领域的很多研究成果(如诗画、唱片、艺术品)很难通过定量指标来彰显其内在价值和对外影响力,此时同行评论等成为优先采用的定性指标,笔者期望未来可以通过图书引文索引、BCI 等指标探讨适用于人文社会科学领域的评价体系。

Altmetrics 是一种评价过滤手段,但不是唯一的评价过滤手段,尽管目前基于 Altmetrics 的研究已经如火如荼地展开,但是其发展得仍然不够成熟,认可度也有待提高,有时还需要通过“公平性测试”等方法进行修正。通过对 PLoS ALMs 上的数据进行分析与研究,作者发现现有工具对 Altmetrics 指标的集成不够完善,大量数据存在缺失等情况,这在一定程度上制约了 Altmetrics 的发展。同时,笔者也注意到在分析论文影响力过程中,不仅应关注 Altmetrics 指标与论文影响力的相关关系,还应该重视不同指标的重要性及量化程度。本文的实证研究发现 Altmetrics 量度与传统计量指标中的引用量度是弱相关关系,在这一具体案例研究中,仍然无法判断 Altmetrics 是否可以替代传统评价方式对论文进行评价。在未来的研究中,笔者将进一步聚焦于提升 Altmetrics 可信度方面,通过多种方法测试并验证方法的有效性。

参考文献:

[1] PRIEM J, TARABORELLI D, GROTH P, et al. Altmetrics: a manifesto[EB/OL]. [2018－05－21]. http://altmetrics.org/manifesto.

[2] LIU J. Metrics and beyond @ SpotOn London 2012[EB/OL]. [2017－11－02]. https://www.altmetric.com/blog/metrics-and-beyond-spoton-london-2012/.

[3] NISO. NISO to develop standards and recommended practices for Altmetrics[EB/OL]. [2017－11－02]. http://www.niso.org/news/pr/view? item_key=72efc1097d4caf7b7b5bdf9c54a165818399ec86.

[4] Proquest. Altmetric data now available in the Summon Discovery Service[EB/OL]. [2017－11－02]. http://www.proquest.com/about/news/2016/Altmetric-data-now-available-in-the-Summon-Discovery-Service.html.

[5] 余厚强, 邱均平. 论替代计量学在图书馆文献服务中的应用[J]. 情报杂志, 2014, 33(9): 163－166.

[6] BRODY T, HARNAD S, CARR L. Earlier Web usage statistics as predictors of later citation impact[J]. Journal of the association for information science & technology, 2006, 57(8): 1060－1072.

[7] WAN J K, HUA P H, ROUSSEAU R, et al. The journal download immediacy index (DII): experiences using a Chinese full-text database[J]. Scientometrics, 2010, 82(3):555－566.

[8] YAN K K, GERSTEIN M. The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics[J]. PLoS one, 2011, 6(5): e19917.

[9] WANG X, LIU C, FANG Z, et al. From attention to citation, what and how does altmetrics work? [EB/OL]. [2017－11－02]. https://arxiv.org/ftp/arxiv/papers/1409/1409.4269.pdf.

[10] ALHOORI H, FURUTA R. Do altmetrics follow the crowd or does the crowd follow altmetrics? [C]//Proceedings of the 14th ACM/IEEE-CS joint conference on digital libraries. Piscataway: IEEE Press, 2014: 375－378.

[11] BORNMANN L. Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics[J]. Scientometrics, 2015, 103(3): 1123－1144.

[12] COSTAS R, ZAHEDI Z, WOUTERS P. Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective[J]. Journal of the Association for Information Science and Technology, 2015, 66(10): 2003－2019.

[13] ZUCCALA A A, VERLEYSEN F T, CORNACCHIA R, et al. Altmetrics for the humanities: comparing goodreads reader ratings with citations to history books[J]. Aslib journal of information management, 2015, 67(3): 320－336.

[14] CHEN K, TANG M, WANG C, et al. Exploring alternative metrics of scholarly performance in the social sciences and humanities in Taiwan[J]. Scientometrics, 2015, 102(1): 97－112.

[15] THELWALL M, HAUSTEIN S, LARIVIÈRE V, et al. Do altmetrics work? Twitter and ten other social web services[J]. PloS ONE, 2013, 8(5): e64841.

[16] PRIEM J, PIWOWAR H A, HEMMINGER B M. Altmetrics in the wild: using social media to explore scholarly impact[EB/OL]. [2017－11－02]. https://arxiv.org/abs/1203.4745.

[17] LIN J. A case study in anti-gaming mechanisms for altmetrics: PLoS ALMs and DataTrust[EB/OL]. [2017－11－02]. http://altmetrics.org/altmetrics12/lin/.

[18] 余厚强. 替代计量指标的换算[EB/OL]. [2017－11－02]. http://blog.sciencenet.cn/blog-441629-893201.html.

[19] 刘春丽. 基于 PLOS API 的论文影响力选择性计量指标研究[J]. 图书情报工作, 2013, 57(7): 89－95.

[20] 顾立平. 论文级别计量研究: 应用案例分析[J]. 现代图书情报技术, 2013(11): 1－7.

[21] 邱均平, 余厚强. 替代计量学的提出过程与研究进展[J]. 图

书情报工作, 2013, 57(19): 5 - 12.
[22] 由庆斌, 汤珊红. 补充计量学及应用前景[J]. 情报理论与实践, 2013, 36(12): 6 - 10.
[23] 刘春丽. 论文层面计量学(Article-Level Metrics):发展过程、特点、指标与应用[J]. 图书馆杂志, 2016, 35(2):63 - 69.
[24] NEYLON C, WU S. Article-level metrics and the evolution of scientific impact[J]. Plos biology, 2009, 7(11):e1000242.
[25] PLOS. Article-Level Metrics[EB/OL]. [2018 - 05 - 21]. https://www.plos.org/article-level-metrics.
[26] 宋玲玲. 基于 PLoS ALMs 的论文影响力评价指标研究——以航空航天医学为例[J]. 情报理论与实践, 2016, 39(9):30 - 36.
[27] 王睿, 胡文静, 郭玮. 常用 Altmetrics 工具比较[J]. 现代图书情报技术, 2015, 30(12): 18 - 26.
[28] 刘晓娟, 宋婉姿. 基于 PLOS ALM 的 altmetrics 指标可用性分析[J]. 图书情报工作, 2016, 60(4):93 - 101.
[29] LAPINSKI S, PIWOWAR H, PRIEM J. Riding the crest of the altmetrics wave: how librarians can help prepare faculty for the next generation of research impact metrics[J]. College & research libraries news, 2013, 119(18):3359 - 3366.
[30] 王雯霞, 刘春丽. 不同学科间论文影响力评价指标模型的差异性研究[J]. 图书情报工作, 2017, 61(13):108 - 116.
[31] PIWOWAR H, PRIEM J. The power of altmetrics on a CV[J]. Bulletin of the American society for information science & technology, 2013, 39(4):10 - 13.
[32] RODGERS E, BARBROW S. A look at altmetrics and its growing significance to research libraries[EB/OL]. [2017 - 11 - 02]. https://deepblue.lib.umich.edu/handle/2027.42/99709.
[33] NIEVES G, MARIA-ISABEL D, ANTONIO C. What role do librarians play in altmetrics? [EB/OL]. [2017 - 11 - 02]. http://eprints.rclis.org/25481/1/Altmetrics.pdf.
[34] RADICCHI F, CASTELLANO C. Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts[J]. Journal of Informetrics, 2012, 6(1): 121 - 130.
[35] LEYDESDORFF L, RADICCHI F, BORNMANN L, et al. Field-normalized impact factors (IFs): a comparison of rescaling and fractionally counted IFs[J]. Journal of the Association for Information Science and Technology, 2013, 64(11): 2299 - 2309.
[36] 陈福佑, 杨立英, 丁洁兰. 不同学科期刊学术影响力比较的方法与实证研究[J]. 图书情报工作, 2013, 57(23):85 - 89.
[37] 王睿, 胡文静, 郭玮. 高 Altmetrics 指标科技论文学术影响力研究[J]. 图书情报工作, 2014, 58(21):92 - 98.
[38] PLOS. Public Library of Science (PLOS) citationInfo [EB/OL]. [2018 - 05 - 22]. http://www.lagotto.io/plos/#citationInfo.
[39] Altmetric support. How is the altmetric attention score calculated? [EB/OL]. [2017 - 11 - 02]. https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-.

作者贡献说明:

刘红煦:提出研究主题,撰写并修改论文;
王铮:提出修改意见。

Study on Altmetrics Academic Quality Evaluation Method Based on Fairness Test

Liu Hongxu[1] Wang Zheng[2,3]
[1] Tongji University Library, Shanghai 200092
[2] School of Public Administration, Northwest University, Xi'an 710127
[3] School of Information Management, Nanjing University, Nanjing 210023

Abstract: [**Purpose/significance**] In order to explore whether the measurement results of Altmetrics indicators are consistent in the evaluation of subject papers, this paper attempts to build a comprehensive literature evaluation model based on Altmetrics through the existing measurement data, and to create new metrics for the evaluation and reuse of data. [**Method/process**] In this study, the Altmetrics indicators of all articles published in the Civil Engineering disciplines from 2009 to 2016 on the PLoS ALMs are obtained. The 15 original indicators are divided into 4 principal components using correlation analysis and principal component analysis methods, which are applied to specific disciplines. It analyzes the correlation between the single index and the comprehensive evaluation index of Altmetrics, and uses the fairness test method to explore the differences in the relevance of time to the evaluation of the paper. [**Result/conclusion**] The study finds that after using the fairness test method to eliminate time impact, the relevance of the citation indicators has increased, while the indicators related to social media communication have declined. But the Twitter index has not decreased, and it has always been highly correlated with the F value.

Keywords: Altmetrics PCA fairness test discipline evaluation